

第一次作业你的报告题目

Code ▼

谢宝进

2024-11-01

- 1 你的主要发现
- 2 数据介绍
- 3 数据概览
- 4 探索性分析
 - 4.1 变量1: 房价的数值描述与图形
 - 4.2 变量2:区域内房价的数值描述与图形
 - 4.3 探索问题1: 房价与建筑面积的关系
 - 4.4 探索问题2:卧室数量与建筑面积的关系
 - 4.5 探索问题3:不同区域的房价差异
- 5 发现总结

1 你的主要发现

1. 房价与区域位置呈现相关性
2. 卧室数量与建筑面积无相关性
3. 房价与建筑面积呈现相关性

2 数据介绍

本报告链家数据获取方式如下:

报告人在2023年9月12日获取了链家武汉二手房网站 (<https://wh.lianjia.com/ershoufang/>)数据。

- 链家二手房网站默认显示100页, 每页30套房产, 因此本数据包括3000套房产信息;
- 数据包括了页面可见部分的文本信息, 具体字段及说明见作业说明。

说明: 数据仅用于教学; 由于不清楚链家数据的展示规则, 因此数据可能并不是武汉二手房市场的随机抽样, 结论很可能有很大的偏差, 甚至可能是错误的。

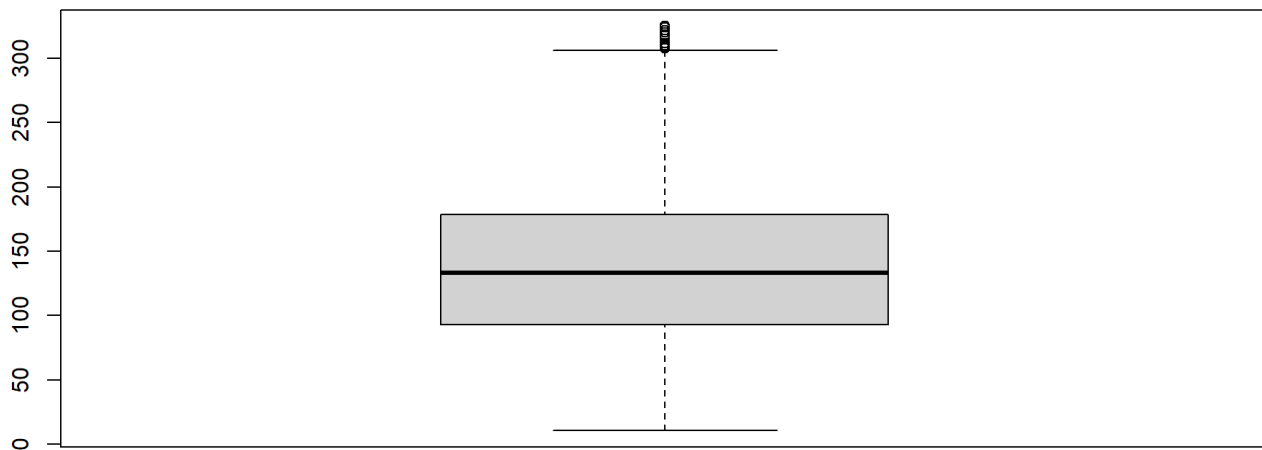
3 数据概览

查看数据表 (lj)结构, 表的前10行示例如下:

```
## spc_tbl_ [3,000 × 18] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ property_name      : chr [1:3000] "南湖名都A区" "万科紫悦湾" "东立国际" "新都汇" ...
## $ property_region    : chr [1:3000] "南湖沃尔玛" "光谷东" "二七" "光谷广场" ...
## $ price_ttl          : num [1:3000] 237 127 75 188 182 ...
## $ price_sqm          : num [1:3000] 18709 14613 15968 15702 17509 ...
## $ bedrooms          : num [1:3000] 3 3 1 3 3 3 2 3 4 3 ...
## $ livingrooms        : num [1:3000] 1 2 1 2 2 2 1 2 1 2 ...
## $ building_area      : num [1:3000] 126.7 86.9 47 119.7 104 ...
## $ directions1        : chr [1:3000] "南" "南" "南" "北" ...
## $ directions2        : chr [1:3000] "北" NA NA "东" ...
## $ decoration         : chr [1:3000] "精装" "精装" "简装" "精装" ...
## $ property_t_height  : num [1:3000] 17 28 18 32 34 34 7 34 5 7 ...
## $ property_height    : chr [1:3000] "中" "中" "低" "高" ...
## $ property_style     : chr [1:3000] "塔楼" "板楼" "塔楼" "塔楼" ...
## $ followers          : num [1:3000] 3 1 3 2 3 1 0 0 2 0 ...
## $ near_subway        : chr [1:3000] "近地铁" NA "近地铁" "近地铁" ...
## $ if_2y              : chr [1:3000] NA "房本满两年" NA "房本满两年" ...
## $ has_key            : chr [1:3000] "随时看房" "随时看房" "随时看房" "随时看房" ...
## $ vr                : chr [1:3000] NA "VR看装修" NA NA ...
## - attr(*, "spec")=
## .. cols(
## ..   property_name = col_character(),
## ..   property_region = col_character(),
## ..   price_ttl = col_double(),
## ..   price_sqm = col_double(),
## ..   bedrooms = col_double(),
## ..   livingrooms = col_double(),
## ..   building_area = col_double(),
## ..   directions1 = col_character(),
## ..   directions2 = col_character(),
## ..   decoration = col_character(),
## ..   property_t_height = col_double(),
## ..   property_height = col_character(),
## ..   property_style = col_character(),
## ..   followers = col_double(),
## ..   near_subway = col_character(),
## ..   if_2y = col_character(),
## ..   has_key = col_character(),
## ..   vr = col_character()
## .. )
## - attr(*, "problems")=<externalptr>
```

```
## [1] 10
```

#极端值处理，并将处理完的数据存储于lj1中



各变量的简短统计：

```

## property_name      property_region      price_ttl      price_sqm
## Length:2856      Length:2856      Min.   : 10.6      Min.   : 1771
## Class :character  Class :character  1st Qu.: 93.0      1st Qu.:10564
## Mode  :character  Mode  :character  Median :133.0      Median :14022
##                                     Mean  :140.9      Mean  :14433
##                                     3rd Qu.:178.2      3rd Qu.:17650
##                                     Max.   :325.0      Max.   :44656
## bedrooms          livingrooms      building_area      directions1
## Min.   :1.000      Min.   :0.000      Min.   : 22.77      Length:2856
## 1st Qu.:2.000      1st Qu.:1.000      1st Qu.: 84.11      Class :character
## Median :3.000      Median :2.000      Median : 94.53      Mode  :character
## Mean   :2.648      Mean   :1.699      Mean   : 97.98
## 3rd Qu.:3.000      3rd Qu.:2.000      3rd Qu.:114.79
## Max.   :7.000      Max.   :4.000      Max.   :230.91
## directions2        decoration      property_t_height  property_height
## Length:2856      Length:2856      Min.   : 2.00      Length:2856
## Class :character  Class :character  1st Qu.:11.00      Class :character
## Mode  :character  Mode  :character  Median :27.00      Mode  :character
##                                     Mean   :24.14
##                                     3rd Qu.:33.00
##                                     Max.   :62.00
## property_style      followers      near_subway      if_2y
## Length:2856      Min.   : 0.000      Length:2856      Length:2856
## Class :character  1st Qu.: 1.000      Class :character  Class :character
## Mode  :character  Median : 2.000      Mode  :character  Mode  :character
##                                     Mean   : 6.316
##                                     3rd Qu.: 6.000
##                                     Max.   :262.000
## has_key            vr
## Length:2856      Length:2856
## Class :character  Class :character
## Mode  :character  Mode  :character
##
##
##

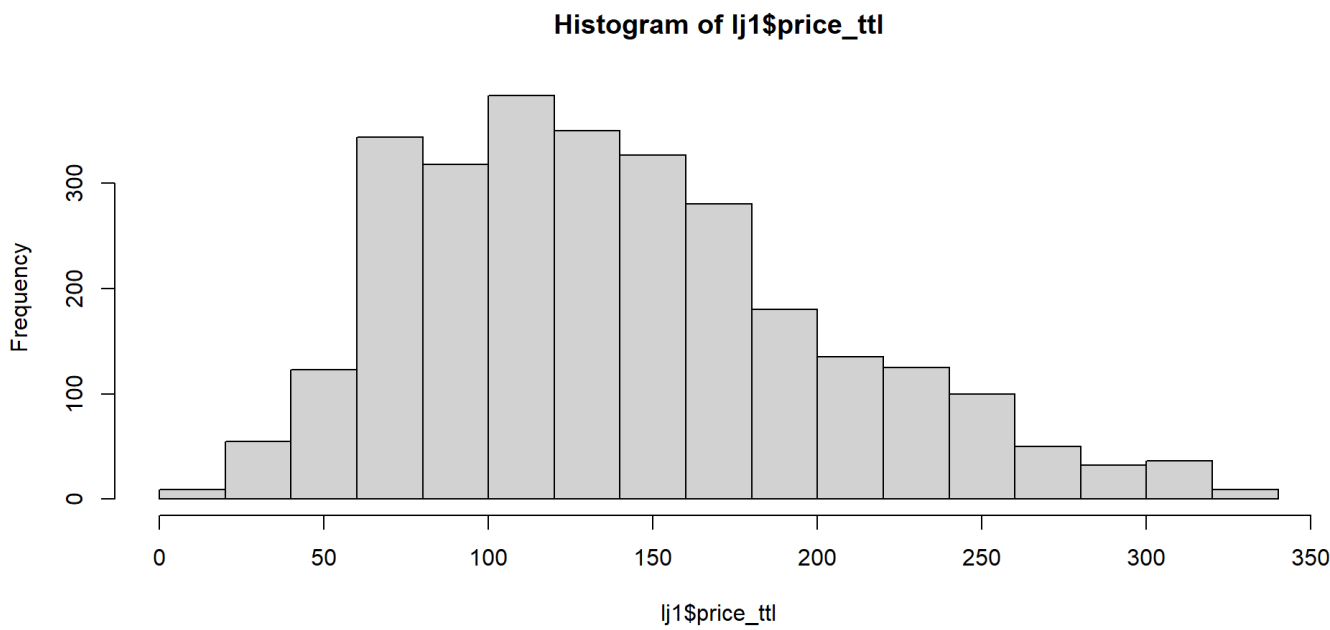
```

可以看到：

- 直观结论1：变量的类型
- 直观结论2：数值型变量的最小值、最大值、平均值、中位值、第一四分位数和第三四分位数

4 探索性分析

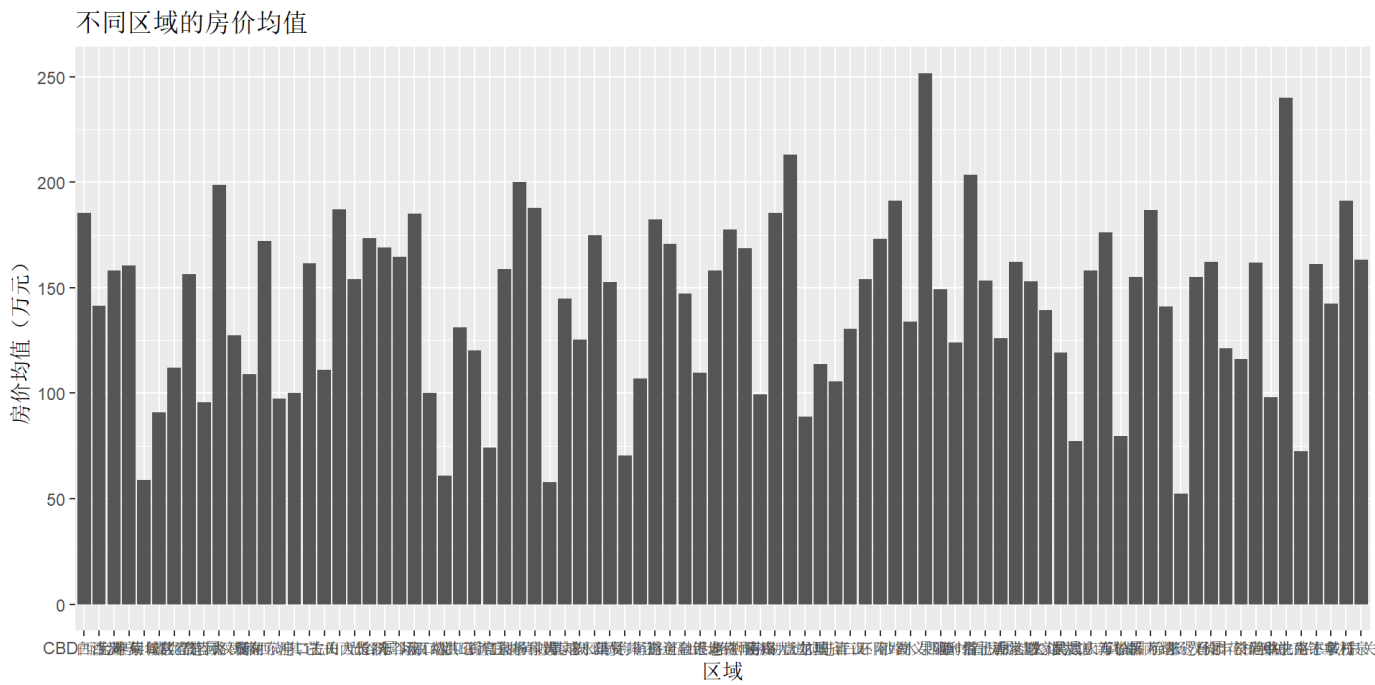
4.1 变量1： 房价的数值描述与图形



发现：

- 发现1:去除极端值后的房价区间为10.6~325
- 发现2： 房价中位数为137

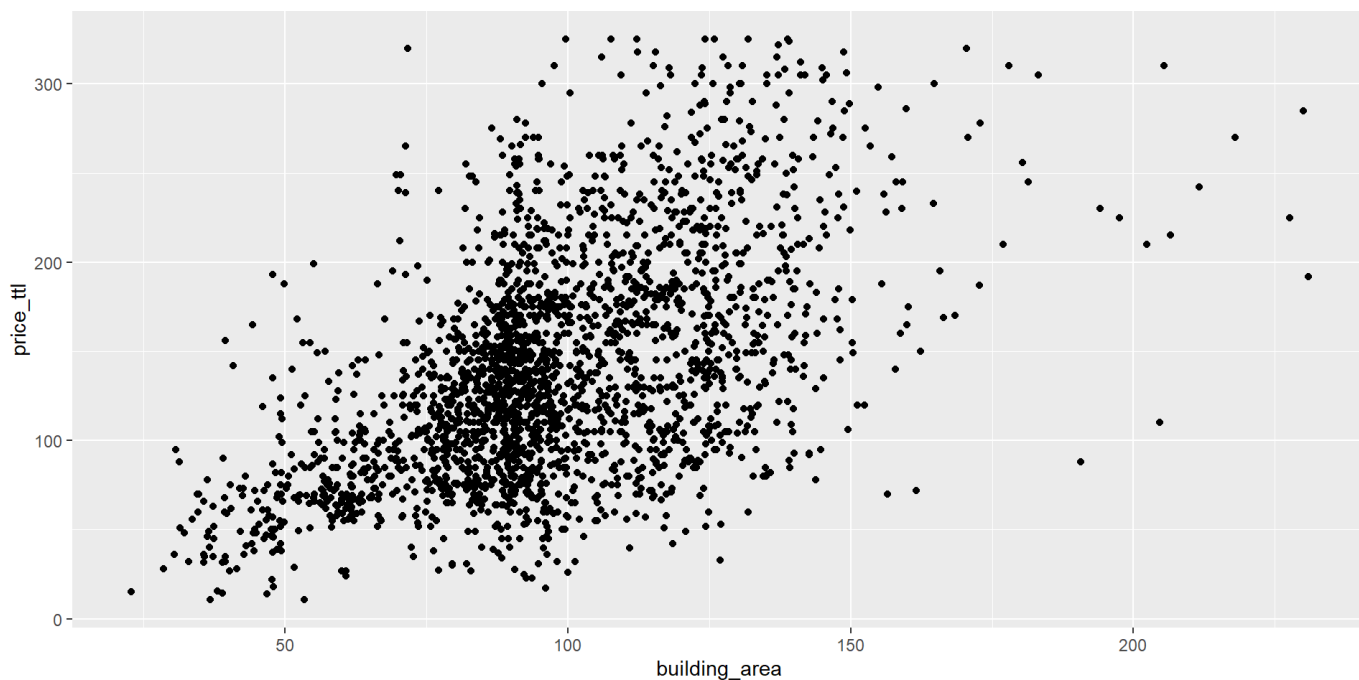
4.2 变量2:区域内房价的数值描述与图形



发现：

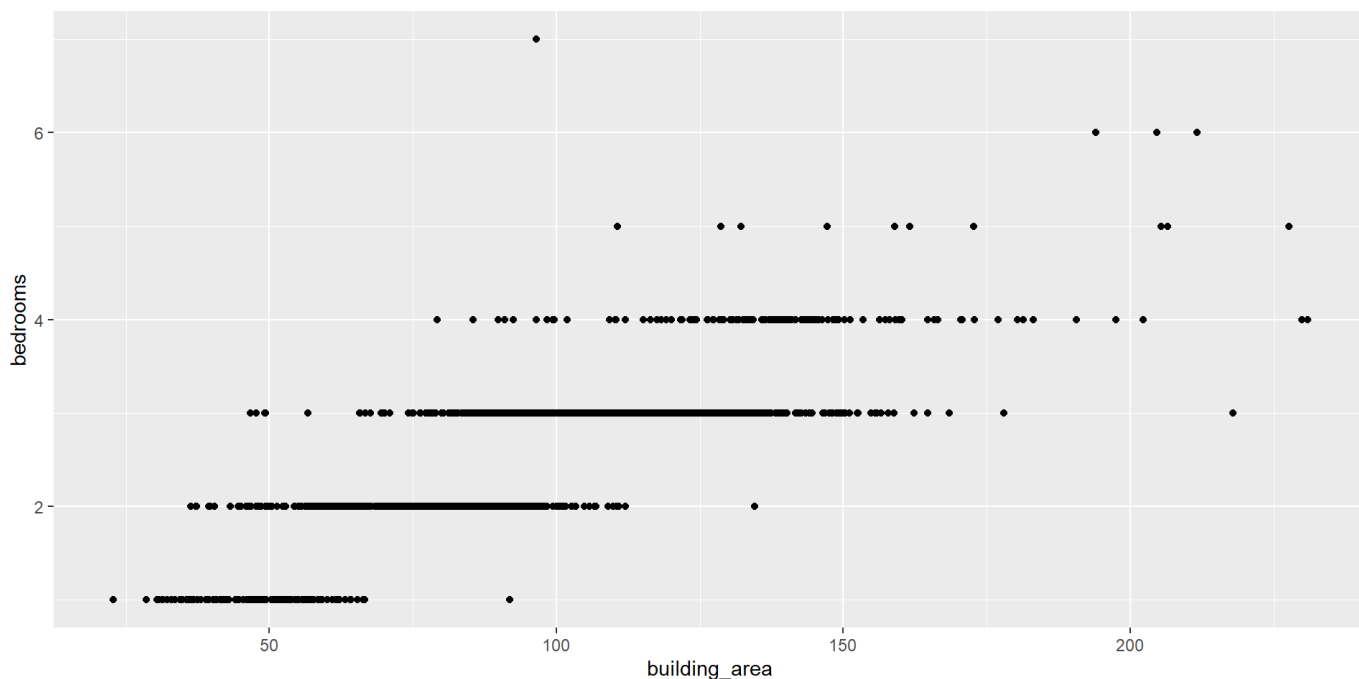
- 发现1:房价与区域位置存在一定关系
- 发现2： 商业区的房价普遍较高

4.3 探索问题1： 房价与建筑面积的关系



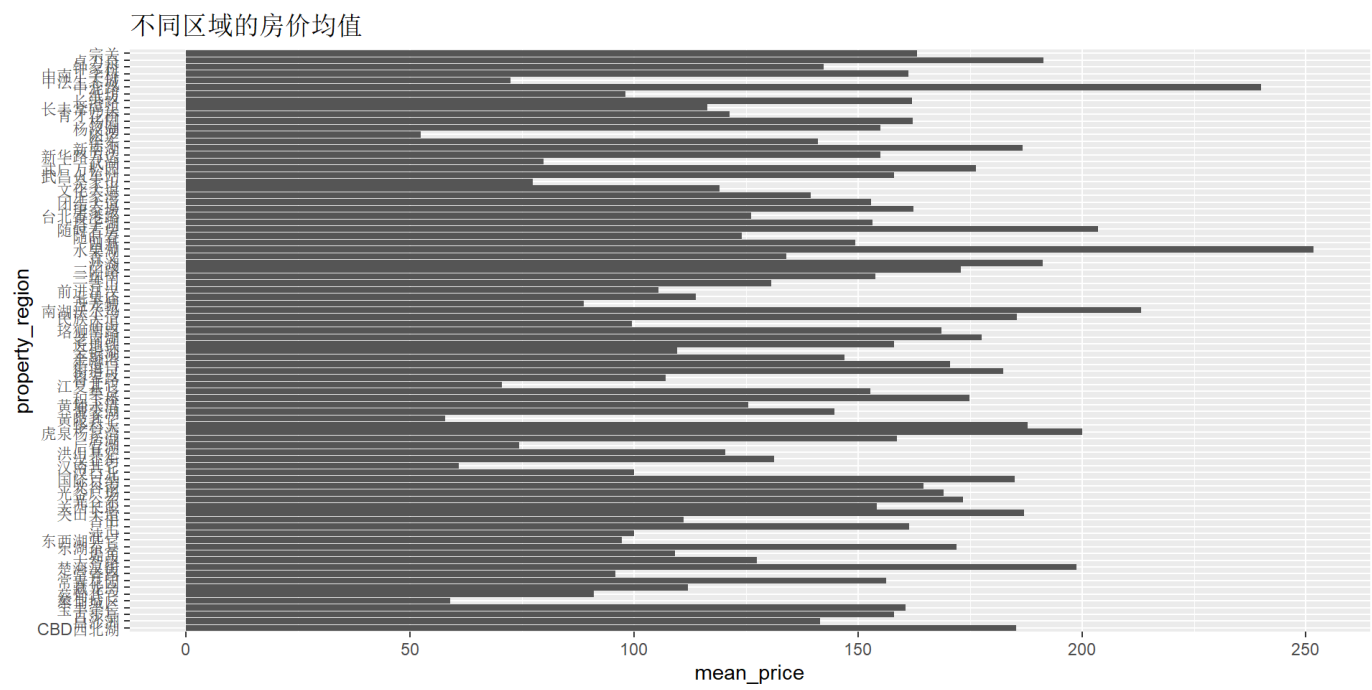
发现： 房价与建筑面积之间也存在正相关关系，建筑面积越大，房价越高。

4.4 探索问题2:卧室数量与建筑面积的关系



发现： 卧室数量与建筑面积之间没有明显的线性关系

4.5 探索问题3:不同区域的房价差异



发现：不同区域的房价存在明显差异，一些区域的房价较高，如 CBD 西北湖、王家湾等，而一些区域的房价相对较低。 —

5 发现总结

- 总结1：商业地段及城市中心地段的房价较高，且均价相差不大。
- 总结2：中心区域价格普遍高于远城区房价
- 总结3：建筑面积越大，房价越高